APPENDIX F

IMPROVEMENTS IN MINIMIZING WET TEST VARIABILITY BY THE STATE OF NORTH CAROLINA

This page intentionally left blank.

IMPROVEMENTS IN MINIMIZING WET TEST VARIABILITY BY THE STATE OF NORTH CAROLINA

F.1 Background

The North Carolina Division of Water Quality (NC DWQ) began in-house WET testing in the late 1970s. Data collected through the mid-1980s indicate that one in four NC NPDES facility effluents tested had the potential to cause acute toxicity instream during low stream flow/high effluent flow conditions (Eagleson et al. 1986). The Division began to require WET self-monitoring by individual facilities in 1985 through administrative letters. DWQ first implemented WET limits in NPDES permits in 1987. As of March 29, 2000, 554 facilities are required to perform some type of WET monitoring; 453 of these have limits. North Carolina permittees have demonstrated compliance rates consistently above 90 percent since the additional TAC were implemented. Chronic *Ceriodaphnia dubia*, acute *C. dubia*, and acute fathead minnow are the primary test types used.

The Division uses two primary strategies to enhance data quality: (1) individual report review and (2) laboratory certification.

Division personnel review each analysis report for the following test acceptability criteria:

- Sample type (specified by permit)
- Sample hold time
- Sample temperature upon receipt at lab
- Control treatment water pH and dissolved oxygen
- Control water hardness*
- Effluent treatment dissolved oxygen
- Test type (specified by permit)
- Replication
- Effluent dilution (specified by permit)
- Control survival and/or reproduction
- Percentage of control organisms producing three broods (*Ceriodaphnia* chronic)
- Control organism reproduction coefficient of variation (*Ceriodaphnia* chronic)*
- Test duration

*NC State criteria

The reviewer may also statistically analyze data sets when the result is unclear based on a cursory review of the data.

The Division's Water Quality Rules specify that WET analyses associated with NPDES permits must be performed by certified laboratories. The Division implemented the laboratory certification program in 1988. Key requirements of that program are specific qualifications for laboratory supervisors, a reference toxicant testing program, annual inspections and audits, and performance evaluation analyses.

Laboratory Supervisor Qualifications

Laboratory supervisors must have either a Bachelor of Science degree in biology or a closely related field and three years of experience in aquatic toxicity testing, or a Master of Science degree in biology or a closely related field and one year of experience in aquatic toxicity testing.

Reference Toxicant Testing Program

The laboratory must maintain a reference toxicant testing program for each organism and test type category (chronic and acute). A reference toxicant test should be performed every two weeks for each organism used in acute WET testing. Alternatively, acute reference toxicant tests may be performed such that NC NPDES acute tests are performed within one week of an acute reference toxicant test for the organism in question. Similarly, a reference toxicant test should be performed once per month for each organism used in chronic WET testing. Alternatively, tests may be performed such that NC NPDES chronic tests are performed within two weeks of a chronic reference toxicant test. To maintain certification for an organism, reference toxicant tests must be performed at least quarterly.

Annual Inspection and Audit

The Division conducts at least one inspection per year at each laboratory. Most inspections are announced, but may be performed without notice. Inspections include the following activities:

- Inspect facilities, equipment, and QA procedures according to the laboratory's standard operating procedures
- Examine living and preserved test organisms
- Review reference toxicant testing program documentation
- Inspect meters and meter calibration records
- Trace randomly selected test records

Performance Evaluation Analyses

The Division may distribute unknown samples to laboratories up to three times per year for analysis. The Division constructs acceptability criteria using the pooled results of the analyses. Laboratories generating results outside of the acceptable range must repeat the analysis. Two consecutive out-of-range results result in decertification. A decertified laboratory regains certification by generating acceptable results on two follow-up analyses.

F.2 Data Evaluation (1992-94) Summary

In January 1992, NC DWQ began recording reproduction data from *Ceriodaphnia* chronic pass/fail tests performed by NC DWQ-certified laboratories in association with NPDES permit requirements. The majority of NC facilities with WET limits use this test. NC pass/fail tests consist of two treatments: a control and a critical concentration, each with 12 replicates. The purposes of the data base were to evaluate the sensitivity of the analysis, assess performance characteristics of the analyses, and evaluate performance of individual laboratories. Analysis was limited to test results with normally-distributed reproduction data.

In 1994, NC DWQ investigators reviewed the PMSD and MSD as a percentage of the control mean for each test (Rosebrock et al. 1994). Evaluation of the data indicated a correlation between PMSD and timing of test termination. EPA methods allow the test to be terminated once 60 percent of the control organisms produce three broods. Therefore, the percentage of adults producing a third brood at test termination may

vary from 60 to 100 percent. Plotting PMSD versus percent of control organisms producing three broods clearly showed that higher percentages of control organisms producing three broods were associated with lower PMSDs (Figure F-1).

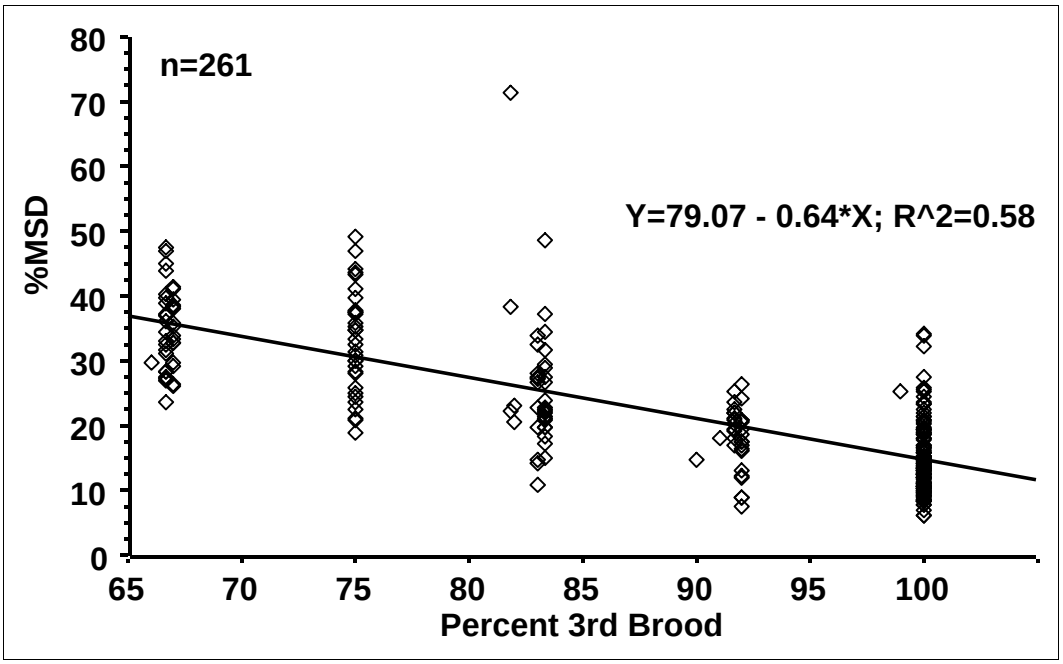

Figure F-1. PMSD versus percent control organisms producing three broods (1994).

Percentile analysis of the PMSD data produced a median PMSD of 20. This means that the "average" analysis, defined as the median, can statistically detect as small as a 20 percent difference between the treatment and control organism reproduction.

Percentile analysis of the CV data for control organism reproduction produced a median of 17 percent and a 95th percentile of 40 percent. This means that 95 percent of the control data sets produced CVs at or below 40 percent.

F.3 North Carolina Chronic Protocol Modifications

Using results from the data evaluations described above and empirical knowledge gained from experience with the test, NC DWQ made several changes to its chronic *Ceriodaphnia* protocol to improve sensitivity, precision, and practical application of test results in its compliance program. These changes were implemented in two stages in late 1994 and early 1996.

December 1994 Changes

- Exclusion of 4th brood and higher neonates from the reproduction analysis
- Addition of a TAC requiring that at least 80 percent of the control organisms produce three broods
- Addition of a TAC requiring that the test be terminated no later than seven days after initiation

January 1996

- Addition of a TAC requiring that the control organism reproduction CV be less than 40 percent
- Specification that for an effluent treatment to be considered as producing an effect, the reproduction mean must be statistically significantly lower than the control mean **and** represent at least a 20 percent reduction from the mean

Reducing the CV of the control reproduction can be shown mathematically to result in reductions in the MSD and PMSD, producing a more sensitive test. Placing an upper limit on the CV will eliminate less sensitive tests. Excluding 4th brood neonates from the reproduction analysis and requiring that at least 80 percent of the control organisms produce a 3rd brood will reduce the control organism reproduction CV.

The specification of at least a 20-percent reduction in reproduction from the control effectively sets a lower limit on test sensitivity. DWQ's experience has shown that high-quality laboratories can produce extremely sensitive tests that can detect very small differences between treatment and control reproduction. Unfortunately, this can be a disincentive for laboratories to produce high quality tests because some clients will gravitate toward laboratories that produce compliant test results. Less sensitive tests will more likely produce such results.

F.4 Evaluation of Program Modifications

The North Carolina data base affords the opportunity to evaluate the effectiveness of additional TAC and changes to the test protocol as they relate to the variability of WET test results. Effluent data for individual laboratories, and across all tests and laboratories, were examined to discern the impact of program changes on laboratory performance. Data were partitioned into two data bases, one for effluent tests completed before December 1994 (termed Pre-1995) and one for effluent tests completed after January 1996 (termed Post-1995). Pass/Fail tests were included in the evaluation. Only tests that did not have a significant mortality effect were considered. Two measures of laboratory performance were calculated using the reproductive data from the tests: PMSD and control CV. The PMSD data set contains all tests reported for compliance. The control CV data set contains all unique controls that were reported by the laboratories and used in compliance calculations. Conclusions reflect the cumulative impact of all changes made to the program from late 1994 to early 1996.

F.5 Overall Test Performance

Pre-1995 and Post-1995 percentile values were generated for the PMSD and the control CV combined across all tests and laboratories (Table F-1). For the PMSD, the median value decreased from 21 percent to 16 percent and the 90th percentile from 39 percent to 31 percent, indicating an overall increase in test sensitivity. The narrower interquartile range of Post-1995 PMSD values (IQR=12 percent), compared with the interquartile range of Pre-1995 PMSD (IQR=16 percent), implies an improvement in the ability of laboratories to achieve similar levels of test sensitivity. (The interquartile range is the difference between the 75th and 25th percentiles of the cumulative distribution function and is a measure of spread of the distribution.) For the control CV, the median value was reduced from 15 percent to 13 percent and the 90th percentile from 34 percent to 28 percent. The overall decrease in the control CV reflects the capacity of laboratories to improve their performance as measured by a decrease in control variability relative to the control mean. Changes in test acceptability criteria and in test protocols improved the consistency of control performance quantified by the reduction in the interquartile range of the control CV Pre-1995 (IQR=15 percent) and Post-1995 (IQR=10 percent).

Table F-1. PMSD and Control Organism CV

	PMSD		CV	
	Pre 1995	Post 1995	Pre 1995	Post 1995
# Tests	4110	5471	2478	3401
Min	0.055	0.049	0.033	0.034
Max	0.839	0.676	0.835	0.400
Median	0.212	0.160	0.155	0.133
IQR	0.164	0.118	0.150	0.103
10^{th} Percentile	0.105	0.095	0.078	0.077
25^{th} Percentile	0.142	0.116	0.103	0.097
50^{th} Percentile	0.212	0.160	0.155	0.133
75^{th} Percentile	0.306	0.233	0.253	0.200
90^{th} Percentile	0.391	0.307	0.343	0.285

F.6 Individual Laboratory Performance

Comparison of effluent data across multiple laboratories provides information about the influence of program changes on individual laboratory performance. Data for a laboratory (Lab 1) with low sensitivity were compared to data from a laboratory (Lab 2) with high sensitivity. Pre-1995 and Post-1995 percentile values were generated for the PMSD combined across all tests for each of the two laboratories (Table F-2). The performance of Lab 2, represented by the distribution of PMSD, was essentially the same Pre-1995 and Post-1995. However, the performance of Lab 1 improved, as evidenced by the changes in medians (33 percent to 18 percent), changes in the 90^{th} percentile (46 percent to 32 percent), and the slight decrease in the width of the interquartile range (13 percent to 12 percent). Additionally, the Post-1995 medians for the two laboratories were relatively close (18 percent and 12 percent) percent for Lab 1 and Lab 2, respectively. A comparison of the cumulative distribution functions for each laboratory indicates that performance was more consistent across laboratories after implementing program changes (Figures F-2 and F-3).

Table F-2. Lab 1 versus Lab 2 PMSD

	Pre-1995		Post-1995	
	Lab 1	Lab 2	Lab 1	Lab 2
# Tests	921	545	1424	466
Min	8.8	5.5	6.8	5.5
Max	67.3	48.9	67.6	39.9
Median	33.5	11.7	18.2	12.5
IQR	13.3	5.5	11.9	4.4

The distribution of PMSD values within a laboratory compared to distributions in other laboratories was examined Pre-1995 and Post-1995 (Figures F-4 and F-5). The range in median values across all laboratories Pre-1995 was 12 percent to 36 percent. Post-1995, the range in median values was 10 percent to 27 percent, indicating a decrease in the overall spread among laboratories. The range in PMSD values within a laboratory was 22 percent to 78 percent Pre-1995. The Post-1995 range in PMSD values within a laboratory compared across laboratories was 17 percent to 61 percent, indicating a narrowing of the range of values within a laboratory (Table F-3). A similar comparison was made using the control CV as an indicator of laboratory ability (Figures F-6 and F-7). The median control CV varied across laboratories from 9 percent to 30 percent Pre-1995. Post-1995, the median control CV ranged across laboratories from 9 percent to 26 percent, a slight improvement in the comparability of control CV. The range in control CVs within a laboratory was 21 percent to 79 percent Pre-1995, while the range in control CVs within a laboratory Post-1995 was 17 percent to 36 percent. Overall, laboratories are generating data with more consistency across, as well as within, laboratories after implementing additional TAC and modifications to testing protocols.

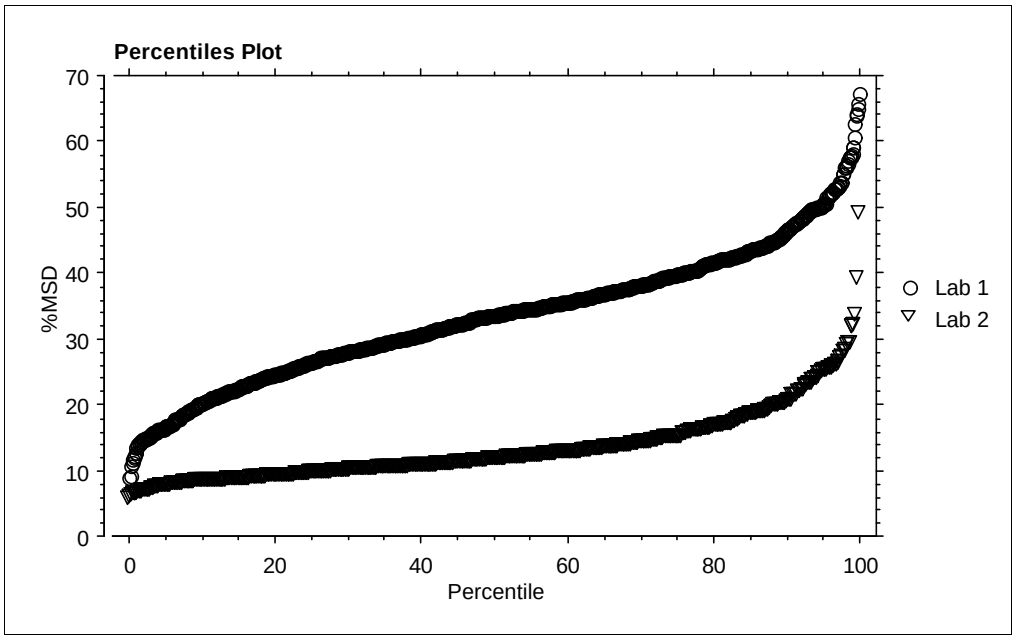

Figure F-2. Laboratory 1 versus Laboratory 2 Pre-1995 PMSD (species: *Ceriodaphnia dubia*).

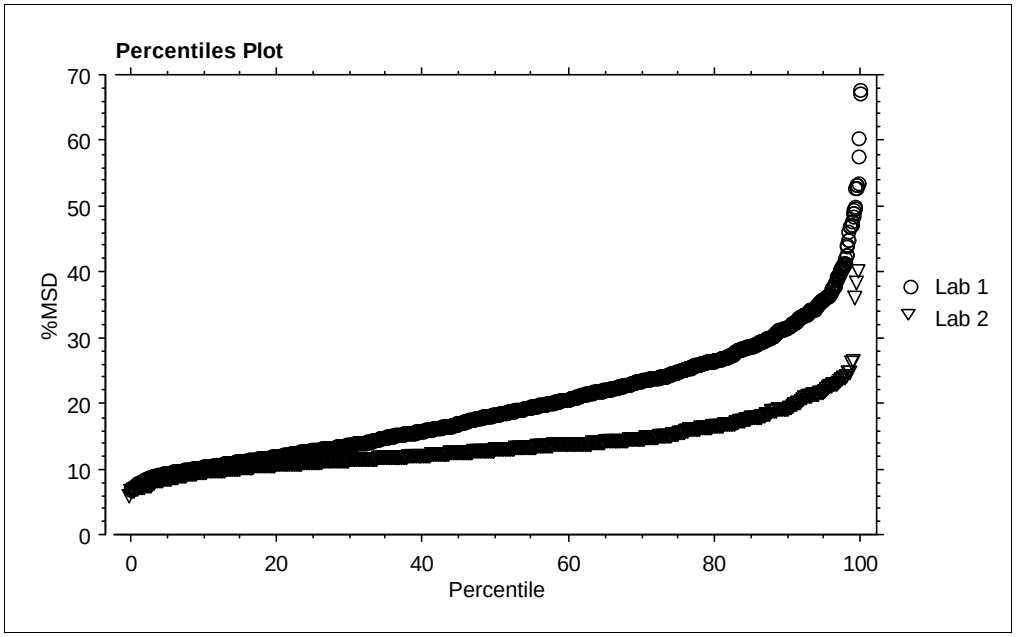

Figure F-3. Laboratory 1 versus Laboratory 2 Post-1995 PMSD (species: *Ceriodaphnia dubia*).

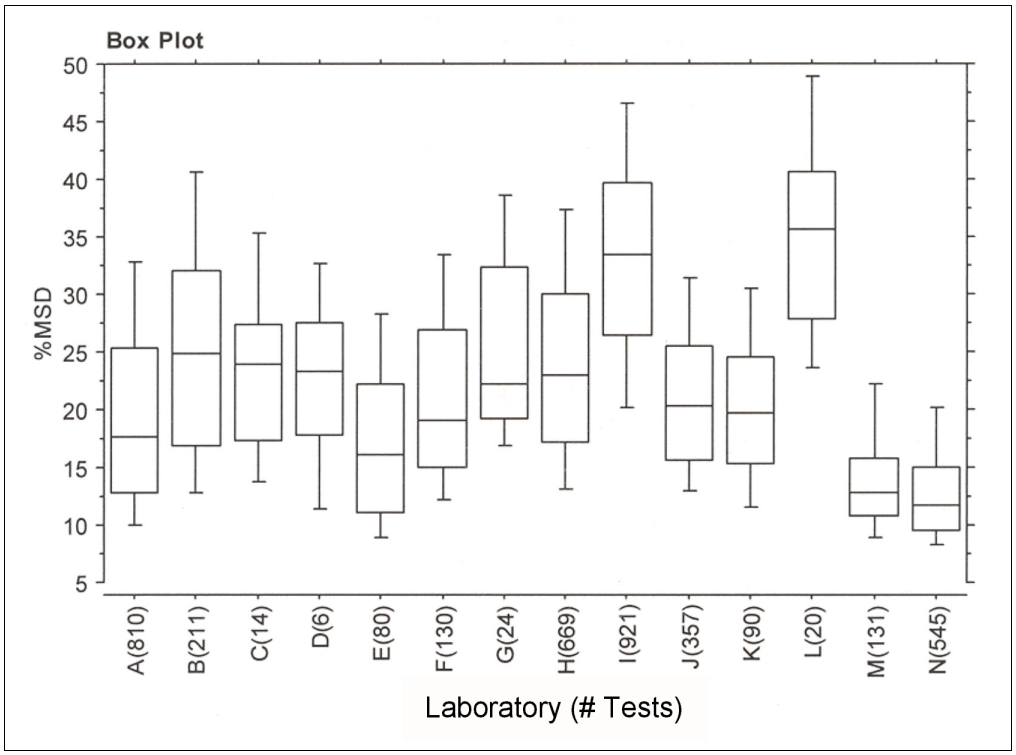

Figure F-4. Individual laboratory performance—Pre-1995 PMSD (species: *Ceriodaphnia dubia*).

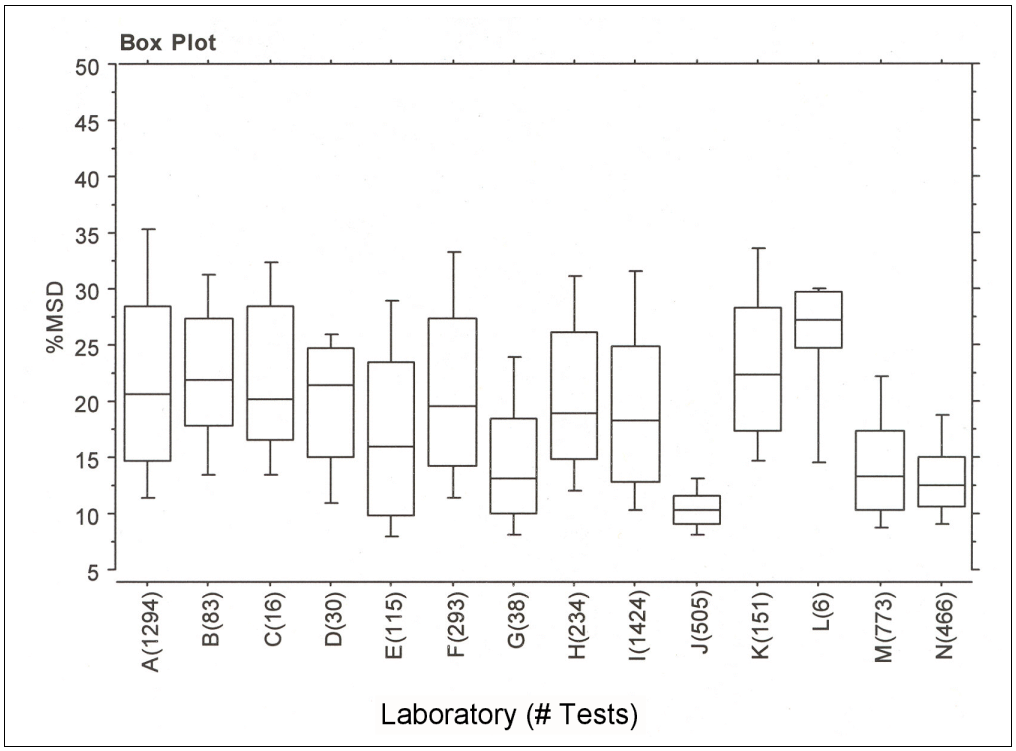

Figure F-5. Individual laboratory performance—Post-1995 PMSD (species: *Ceriodaphnia dubia*).

Table F-3. Descriptive Statistics—PMSD

	Pre-1995						Post-1995					
Lab	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**
A	810	6.0	83.9	77.9	17.6	12.6	1294	6.4	58.9	52.5	20.6	13.7
B	211	8.6	59.7	51.1	24.8	15.0	83	10.2	39.9	29.7	21.9	9.6
C	14	13.7	35.6	21.9	23.9	10.0	16	12.5	34.5	22.1	20.1	11.9
D	6	10.6	33.2	22.6	23.3	9.7	30	9.6	33.9	24.3	21.5	9.6
E	80	6.5	43.5	37.0	16.1	11.1	115	5.6	43.8	38.3	15.9	13.6
F	130	6.9	69.4	62.5	19.1	11.8	293	6.8	55.0	48.2	19.5	13.0
G	24	13.9	45.0	31.1	22.2	13.2	38	6.6	33.1	26.5	13.1	8.4
H	669	6.2	71.5	65.3	23.0	12.8	234	8.4	38.9	30.5	19.0	11.4
I	921	8.8	67.3	58.4	33.5	13.3	1424	6.8	67.6	60.8	18.2	11.9
J	357	8.7	69.8	61.1	20.4	9.7	505	6.4	26.0	19.5	10.2	2.5
K	90	9.7	55.5	45.8	19.7	9.1	151	8.3	47.6	39.3	22.4	10.9
L	20	22.0	59.0	37.0	35.7	12.9	6	13.4	30.1	16.7	27.2	5.0
M	131	6.4	49.9	43.5	12.9	5.0	773	4.9	40.3	35.3	13.3	6.9
N	545	5.5	48.9	43.4	11.7	5.5	466	5.5	39.9	34.4	12.5	4.4

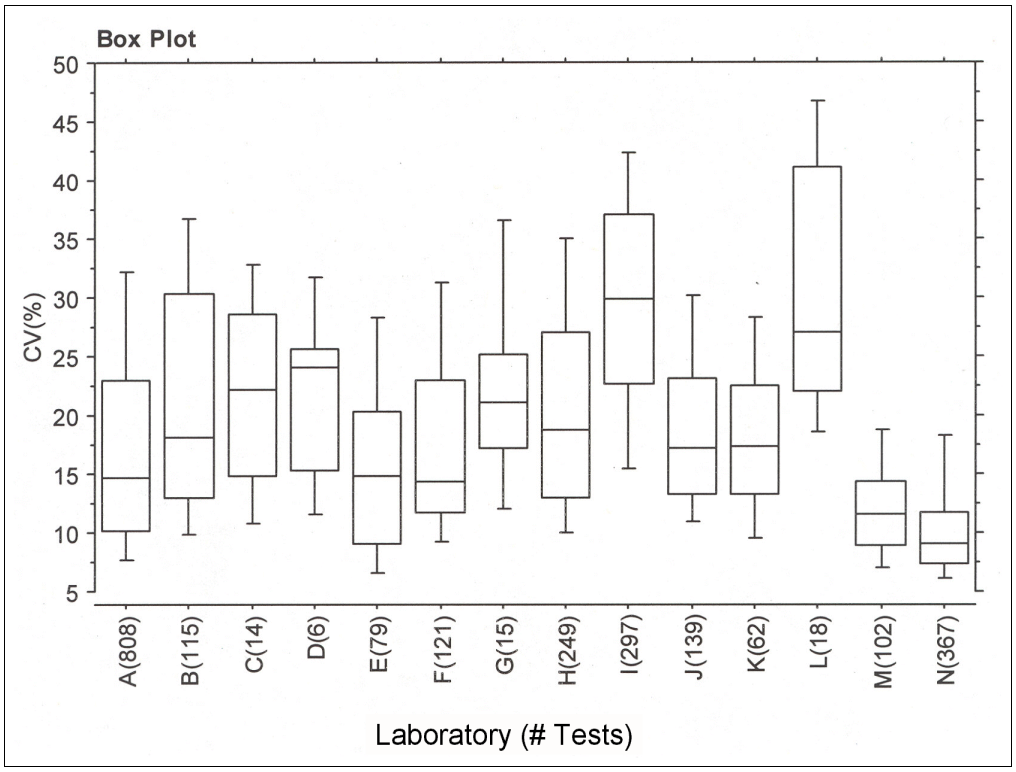

Figure F-6. Individual laboratory performance—Pre-1995 CV (species: *Ceriodaphnia duba*).

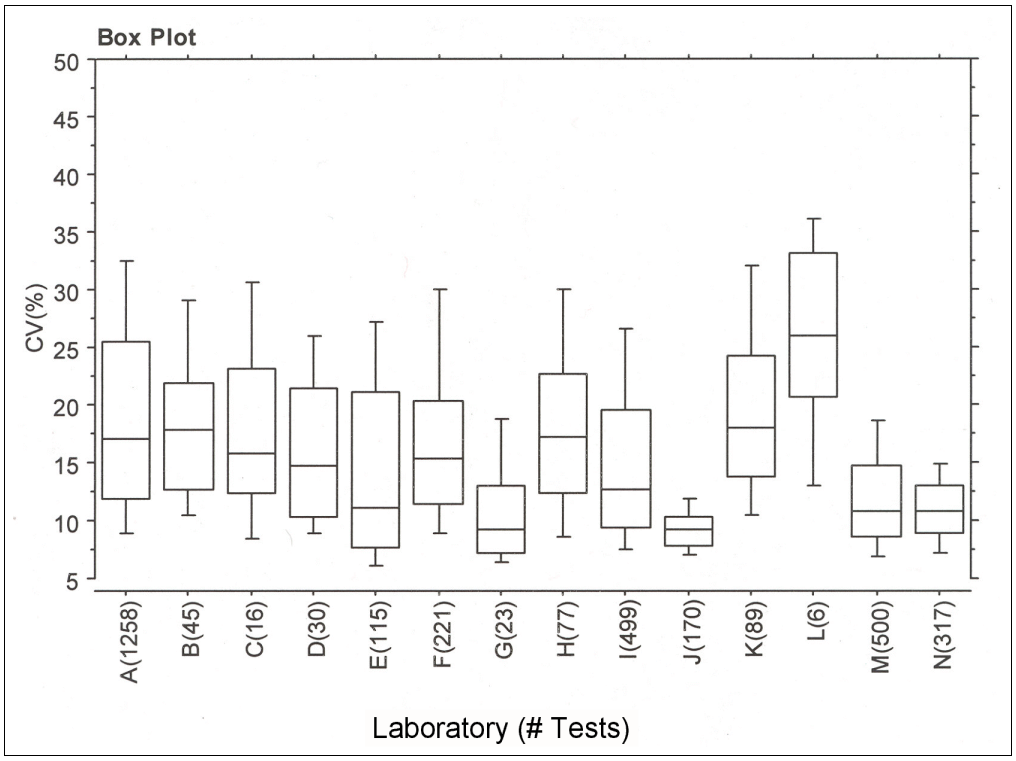

Figure F-7. Individual laboratory performance—Post-1995 CV (species: *Ceriodaphnia dubia*).

Table F-4. Descriptive Statistics—Coefficient of Variation (CV)

	Pre-1995						Post-1995					
Lab	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**
A	808	0.041	0.835	0.794	0.146	0.129	1258	0.043	0.399	0.356	0.171	0.136
B	115	0.062	0.511	0.450	0.182	0.173	45	0.059	0.361	0.302	0.178	0.092
C	14	0.092	0.334	0.242	0.222	0.137	16	0.066	0.378	0.311	0.158	0.109
D	6	0.112	0.324	0.212	0.241	0.102	30	0.074	0.332	0.258	0.147	0.111
E	79	0.041	0.374	0.333	0.148	0.112	115	0.038	0.400	0.362	0.111	0.134
F	121	0.051	0.516	0.464	0.143	0.113	221	0.062	0.384	0.322	0.152	0.090
G	15	0.113	0.404	0.291	0.211	0.080	23	0.050	0.343	0.293	0.092	0.059
H	249	0.055	0.610	0.555	0.188	0.140	77	0.061	0.379	0.318	0.171	0.103
I	297	0.068	0.672	0.604	0.299	0.144	499	0.047	0.399	0.352	0.127	0.101
J	139	0.071	0.596	0.525	0.172	0.098	170	0.054	0.222	0.168	0.092	0.025
K	62	0.046	0.564	0.517	0.173	0.093	89	0.047	0.392	0.345	0.180	0.104
L	18	0.138	0.571	0.433	0.271	0.190	6	0.121	0.365	0.245	0.259	0.124
M	102	0.053	0.398	0.345	0.115	0.056	500	0.034	0.341	0.307	0.107	0.062
N	367	0.033	0.472	0.439	0.091	0.043	317	0.038	0.333	0.296	0.108	0.040

REFERENCES

Eagleson, K.W., S.W. Tedder, and L.W. Ausley. 1986. Strategy for whole effluent toxicity evaluations in North Carolina. In *Aquatic Toxicology and Environmental Fate: Ninth Volume, ASTM STP 921*. T.M. Poston, R Purdy, eds. American Society for Testing and Materials. Philadelphia, PA. 154-160.

Rosebrock, M.M., N.W. Bedwell, and L.W. Ausley. 1994. Indicators of *Ceriodaphnia dubia* chronic toxicity test performance and sensitivity. Poster presentation, Society of Environmental Toxicology and Chemistry 15th Annual Meeting, Denver, CO.